

Automatic parameter tuning for functional regionalisation methods

José Manuel Casado-Díaz¹, Lucas Martínez-Bernabéu², and Francisco Flórez-Revuelta³

¹International Economics Institute and Department of Applied Economic Analysis, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain (e-mail: jmcasado@ua.es)

²International Economics Institute, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain (e-mail: lucas.martinez@ua.es)

³Faculty of Science, Engineering and Computing, Kingston University, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, United Kingdom (e-mail: f.florez@kingston.ac.uk)

Abstract. The methods used to define functional regions for public statistics and policy purposes need to establish several parameter values. This is typically achieved using expert knowledge based on qualitative judgements and lengthy consultations with local stakeholders. We propose to support this process by using an optimisation algorithm to calibrate any regionalisation method by identifying the parameter values that produce the best regionalisation for a given quantitative indicator. The approach is exemplified by using a grid search and a genetic algorithm to configure the official methods employed in the UK and Sweden for the definition of their respective official concepts of local labour markets.

JEL classification: R12, R23, R28, R58

Key words: Labour market areas, functional regions, parameter tuning, regionalisation, travel-to-work areas, algorithm configuration

Running title: Parameter tuning for regionalisation methods

Corresponding author: Lucas Martínez-Bernabéu (lucas.martinez@ua.es)

1 Introduction

A labour market area (LMA) is a type of functional region (FR) that reflects the territorial reality where the supply of and the demand for labour meet at the local level. In practice, the process of identifying the boundaries between one LMA and the rest typically involves the analysis of travel-to-work flows between the territorial units (TUs) that are taken as the building blocks through aggregative procedures that are very diverse in nature. As such, this type of regions differs from those defined as relatively homogeneous in terms of a selected group of characteristics.

FRs and the specific case of LMAs are an object of interest for many public administrations and there are many international examples of them being identified officially (EUROSTAT and Coombes, 1992; OECD, 2002; Casado-Díaz and Coombes, 2011). The goal that is pursued in the vast majority of cases (sometimes explicitly and in others, tacitly) is the definition of LMAs that are highly integrated in internal terms and defined by boundaries that separate them from other similarly cohesive areas, with which their functional links are significantly weaker. The aim is therefore the definition of LMAs composed of TUs that (i) exchange numerous and abundant flows of workers and (ii) that are relatively self-contained with regard to the other LMAs. The differences between the diverse methods used to make this concept operative include many dimensions, such as whether they are based on the initial definition of foci around which LMAs are built,

whether commuting flows are measured in reciprocal terms (from A to B, and from B to A) or only in one direction (from the hinterland towards a given centre, for example), whether the flows are taken in absolute (e.g., “number of workers commuting from A to B”) or relative terms (e.g., “share of workers commuting from A to B” or “share of jobs in A filled by commuters from B”), which specific function is used to “measure” the links, and whether the method consists of the application of a series of definitional rules, separately devised to reveal the underlying patterns, or is based on a single rule that is applied from start to finish (which allows distinguishing between rule-based and hierarchical methods).

What all the methods share, both those that are simpler and the ones involving more complicated steps, is the need to exogenously fix values for a broader or smaller set of parameters. Some examples of well-established official methods can illustrate this. The method used by Statistics Sweden to define the national set of LMAs (Statistics Sweden, 2010) has a first step in which TUs (*kommuner*) are identified as potential centres where (a) the percentage of residents working within the TUs’ boundaries exceeds 80% and in which (b) the maximum outgoing flow to a single destination is less than 7.5% of the working population (a pair of TUs is grouped if they have each other as destinations of their largest commuting outflows). The rest of the TUs composing the Swedish territory are then assigned to the destination of their largest outflow through a hierarchical process until all TUs are allocated to LMAs (all of which must include a centre).

The much more complex method used in Italy (ISTAT, 1997) includes (a) the selection of TUs (municipalities) as potential foci, (b) the consolidation of those potential foci with strong connections, (c) the assignment of other TUs to form proto-LMAs, and (d) the final identification of such LMAs. The different steps include, among others, these parameters: 20% TUs ranking higher in “centrality” or “self-containment” are considered potential foci with a concentration of jobs, two of these potential foci are combined if (i) the commuting flow from A to B exceeds 10% of the total out-commuting from A, (ii) commuting flow from B to A exceeds 1% of the total out-commuting from B, and (iii) the value for the “weighted interaction function” exceeds 0.2. To be considered a LMA, a group of TUs must satisfy a double condition: (a) its size must be over 1,000 jobs and (b) the self-containment level must be at least 75% (at least 75% of the residents must work locally, which is called supply-side self-containment, and at least 75% of the jobs must be filled by local residents, which is called demand-side self-containment).

The ISTAT’s method has many similarities with the one that was used in the UK for the definition of the national version of LMAs, the Travel-to-Work Areas (TTWAs) (Coombes et al., 1986), until the revision that followed the dissemination of data from the Census of Population 2001, when the procedure was largely simplified (Office for National Statistics et al., 2008). TTWAs constitute one of the examples of LMAs with a longer history and have also been one of the concepts more widely applied in international terms (Casado-Díaz and Coombes, 2011). In its newer version, this procedure initially assumes that every TU is a potential LMA. The method then proceeds by iteratively considering the LMA with the lowest score on the criteria of validity. If that LMA does not fulfil the set criteria, it is dismembered, and its constituent TUs are reassigned to whichever remaining potential LMAs that score highest on the so-called aggregation criteria (an interaction index). The criteria of validity are codified in terms of a linear trade-off relationship between self-containment, which is measured as the minimum of both demand- and supply-side self-containment, and population size, which is in terms of economically active population; thus, for a given minimum self-containment level (e.g., 70%), a target population size should be reached (e.g., 25000), and for a given minimum population size (smaller than the target size, e.g., 3500), a target self-containment level (greater than the

minimum level, e.g., 75%) should be reached. The method includes therefore the need for fixing four values: target and minimum values for both self-containment and size (the specific values were 66.7-70% and 3,500-25,000 in the last revision).

What the preceding examples of official methods show is that defining LMAs in a given territory involves selecting not only one of the formal procedures available, or creating a new one (clearly the preferred option given the international experience reviewed in the references already cited), but also selecting specific values for the parameters embedded in those methods. These choices are typically based on what could be called “expert knowledge”, which is used in the extensive experimentation phase that characterises the definition processes, and is modulated by the opinions/suggestions from central and/or regional and/or local authorities, and (less frequently) other relevant stakeholders. Those values are typically updated when LMAs and other related geographies are reconsidered following the availability of new data through what could be considered a trial and error method in which territorial changes and the predominant opinion about how they should be reflected in the new boundaries are considered. Examples of these changes are the modifications of the self-containment goals in the TTWA updates (Office for National Statistics et al., 2008, p. 14) or the changes in the thresholds used to classify a county as part of a *metropolitan area* in the USA, where this type of FR constitute the geography of LMAs in the most populated parts of the country (Office of Management and Budget, 2000).

This study proposes a general methodology aimed at supporting the decision-making process with regard to parameter setting in regionalisation methods and, more specifically, in the procedures used by public administrations for the delineation of their official sets of LMAs.

The structure of this article is as follows. In Section 2, the use of a “wrapper” model is proposed, in which a parameter setting algorithm contains or “wraps” round one specific regionalisation algorithm (see Fig. 1). The parameter setting algorithm—which can be any form of optimisation, such as a grid search or simulated annealing—performs a search for a good set of parameters for the regionalisation algorithm by generating candidate sets of parameters and comparing the fitness values of the regionalisations resulting from those parameters. Section 3 summarises the two fitness functions that have been used in this context and our combined proposal. Section 4 describes the two parameter setting techniques considered in this work: grid-search and genetic algorithm. Section 5 evaluates the proposed methodology by applying the two selected techniques to two different national cases, UK and Sweden, and comparing the resulting regionalisations with their official definitions of LMAs. This section includes an additional illustration of how the approach could be equally applied in a restricted context (e.g., when the allowed range of a specific parameter is fixed a priori to meet absolutely essential statistical or policy-making objectives). Section 6 discusses the results and, to conclude, Section 7 summarises the contribution of this study and the limitations of the approach. Some potential extensions as well as different alternatives are also summarised there.

2 Automating parameter setting in regionalisation algorithms

2.1 The wrapper-based strategy

The approach proposed here is related to the literature on “wrappers”, which was initially developed by John et al. (1994) and applied in fields such as parameter setting and

feature selection in support vector machines, used for classification and regression analysis (Cantú-Paz, 2004; Saeys et al., 2007; Casado Yusta, 2009).

The strategy is based on the combination of two separate algorithms. The parameter-setting algorithm (the “wrapper”) is used to tune the regionalisation algorithm, by identifying the set of regionalisation parameters that allows the regionalisation algorithm to reach the maximum value of a given fitness function. In other words, the wrapper performs a search for the best possible set of parameters for the regionalisation algorithm by generating sets of parameters, applying that regionalisation algorithm using them and evaluating the resulting regionalisations through the chosen fitness function. Fig. 1 illustrates the approach.

2.2 Problem formalisation and notation

The result of a regionalisation algorithm A on a regionalisation problem instance $I = (S, T)$ is a possible partition \mathcal{P} of the base territorial units (TUs) to be divided into LMAs, where $S = \{i, j, \dots\}$ is the set of $N = |S|$ TUs and T is the matrix of commuting flows between such TUs so that T_{ij} is the number of residents in TU i that work in TU j (note that the diagonal of the matrix is not null). For simplicity in the formulation, we will represent the LMA of a given TU i in partition \mathcal{P} as $M^{\mathcal{P}, i}$, the aggregated commuting flow from a LMA X to another LMA Y as $T_{XY} = \sum_{\forall i \in X} \sum_{\forall j \in Y} T_{ij}$, the number of employed residents in LMA Z as $O_Z = \sum_{\forall i \in Z} \sum_{\forall j \in S} T_{ij}$ and the number of jobs as $J_Z = \sum_{\forall i \in Z} \sum_{\forall j \in S} T_{ji}$.

Such an algorithm A has p control parameters, each with an associated domain of values, and the space of possible sets of parameters (called *parameter configurations*), Θ , is the cross-product of these domains (or a subset if some combinations are not allowed).

The problem addressed in this study (which could be termed *parameter setting* or *algorithm configuration problem*) is to find the parameter configuration $\theta \in \Theta$, for a given regionalisation algorithm A and a problem instance I , that produces the result $\mathcal{P} = A(\theta, I)$ with the highest score on a fitness function f :

$$\arg \max_{\theta} f(A(\theta, I)) \quad (1)$$

The choice of the regionalisation algorithm is not part of the problem (each administration/practitioner would be using the method that they consider best), but it is required to select a fitness function and an optimisation method as the parameter “tuner”. For the latter, there are many alternatives, and depending on the problem instance, several of them would be equally suitable to find the optimal parameter setting. However, for the former, there are few references upon which to base a decision. Section 3 summarises the dilemma of choosing a suitable fitness function for the problem whilst Section 4 describes the two alternative methods tested in this work.

3 Choosing a fitness function

The choice of this function is not trivial due to the lack of precedents because the literature that addresses the definition of LMAs and other types of FRs has mainly relied on the use of heuristic, greedy methods that identify a single, relatively good solution, in a short period of time. This type of algorithm does not require a choice between different solutions and therefore a (global) fitness function is not needed. Moreover, there is little information that can be useful for the construction of a fitness function in the few quantitative analyses devoted to the assessment of a given regionalisation that can be found

in the literature, as these studies do not have the objective of analysing the optimality in terms of the main objectives of the ideal definition of a LMA (inner cohesion and self-containment), but have concentrated on their behaviour as homogeneous units in the attribute space (e.g. Cörvers et al., 2009).

In the following two sub-sections, we summarise the two main available fitness functions provided by recent literature, and we propose in sub-section 3.3 an alternative that combines both of them.

3.1 Inner interaction index

Among the scarce examples of the use of fitness functions in this context, two studies stand out (Flórez-Revuelta et al. (2008) and Martínez-Bernabeu et al. (2012)), in which the definition of LMAs is approached as an optimisation problem subject to several conditions. In these articles, different Genetic Algorithms (GAs) are used to maximise a certain fitness function subject to the same restrictions used in the TTWA method. That methodology is presented as an alternative to the greedy approaches that currently dominate and has proven to succeed in detecting more LMAs with similar or higher levels of cohesion and self-containment compared to the TTWA method when the same parameter set is used in both procedures (the preference for detail in this type of exercise is discussed in Casado-Díaz and Coombes, 2011).

Their fitness function, which we will call the *II* function (from *inner interaction*), is the sum for each TU of an interaction index between that TU and the rest of the region (LMA) of which it is part of. This function can be expressed as follows:

$$II^S(\mathcal{P}) = \frac{\sum_{i \in \mathcal{P}} I^S(\{i\}, M^{\mathcal{P},i} \setminus \{i\})}{N} \quad (2)$$

where I^S , an index that was originally proposed by Smart (1974), measures the interaction between two regions X and Y (in this case, between the region composed of TU i , denoted by $\{i\}$ in the equation, and the rest of its LMA). This index was used to guide the grouping of territorial units in the TTWA method (Coombes et al., 1986):

$$I^S(X, Y) = I^S(Y, X) = \frac{T_{XY}^2}{O_X J_Y} + \frac{T_{YX}^2}{O_Y J_X} \quad (3)$$

Eq. 2 includes a normalisation term N (the number of TUs in the territory) that was not part of the original definition. It does not alter the results of using that measure but allows confining the results in the range of [0..1] (experimentation shows that values greater than 0.1 are extremely unlikely) and comparing the results between territories composed of a different number of base TUs.

3.2 Modularity quality index

Based on similar premises, addressing the regionalisation problem as an optimisation one, other authors have presented different methods for the delineation of LMAs. Thus, Fusco and Cagliani (2011) have compared the results from three hierarchical functional regionalisation methods: two forms of polarised (core-based) functional clustering and a form of heuristic modularity optimisation through hierarchical aggregation. The latter is based on the work by Blondel et al. (2008), a fast algorithm for community detection in large networks, but in this case the search process is guided by the Modularity Q index, developed by Newman and Girvan (2004) for the detection of communities in

networks. These communities are groups of nodes with dense connections internally and sparser connections between groups. In the context of this study, nodes correspond to TUs and communities to LMAs. That same fitness function and the original methodology proposed by Newman and Girvan (2004) has also been used by Farmer and Fotheringham (2012) to perform a recursive spectral bi-partitioning in a divisive hierarchical clustering.

Modularity Q is calculated by a comparison between the fraction of the total commuting observed within each community and the expected value of that fraction in a null model: a mirrored network whose nodes have the same degree distribution as the real network (each pair of corresponding nodes have the same number of workers) but with links (commuting flows) that are uniformly distributed among all the nodes. In a weighted directed network, such is our case, it can be formulated as (Leicht and Newman, 2008):

$$Q(\mathcal{P}) = \sum_{M \in \mathcal{P}} \left(\frac{T_{MM}}{O_{\mathcal{P}}} - \frac{O_M J_M}{O_{\mathcal{P}}^2} \right) \quad (4)$$

where the first term of the difference is the actual fraction of commuting comprised within the LMA and the second term is the expected value for such LMA in the corresponding null model. The value of $Q(\mathcal{P})$ ranges from -1 to 1 , where values higher than 0 indicate higher than expected modularity. In the community detection literature values over 0.7 are considered to be an evidence of strongly differentiated communities, but there are no studies in the context of spatial functional regionalisation.

Despite its potential, the use of Modularity Index Q has been criticised in the context of community detection by some authors such as Fortunato and Barthelemy (2007), who warn against a resolution limit, which in large networks would result in an inability to identify the actual communities if their sizes vary too much, as is typical in real world data. Moreover, Lancichinetti and Fortunato (2011) have recently shown that this resolution limit cannot be solved through the introduction of tunable parameters because “modularity suffers from two co-existing problems: the tendency to merge small sub-graphs, which dominates when the resolution is low, and the tendency to split large sub-graphs, which dominates when the resolution is high” (Lancichinetti and Fortunato, 2011, p. 1), even when the actual communities are easily detected by other methods and visual inspection.

3.3 Inner interaction and modularity combined

There is a lack of evidence on the merits of each of the cited fitness functions or any other in the context of functional regionalisation and, specifically, in LMA definition. Our preliminary experimentation with both of these functions has shown that none of them seem to be sufficient as the only objective to maximise in this context.

Thus, the parameters that maximised the interaction-based function resulted in very fine-grained regionalisations with excessively low levels of self-containment—approximately 50%, which is much below the standard in this field according to the literature (Casado-Díaz and Coombes, 2011)—for most of the identified LMAs. This function produced good results when it was used by Flórez-Revuelta et al. (2008) and Martínez-Bernabeu et al. (2012) in a different context, in which the function was maximised subject to certain parameter values set a priori (those used in the UK for their TTWAs). The goal here is substantially different: the identification of the most appropriate set of parameters for a given regionalisation method and problem instance.

With the modularity function, the opposite occurred: the highest modularity was reached for parameters considerably greater than those used in practice, and the resulting

regionalisations include extremely wide areas where the self-containment was very high but the levels of inner interaction were noticeably low. All those results were exceedingly contradictory to what experts have considered appropriate for the past three decades (Casado-Díaz and Coombes, 2011).

A reasonable approach to this issue seems to be defining a fitness function that offers a suitable trade-off between high cohesion levels but low self-containment of the interaction-based (Eq. 2) results and high self-containment but low cohesion of the modularity-based (Eq. 4) results. We suggest achieving this goal by (a) appropriately transforming the interaction index (Eq. 3) and/or (b) combining the resulting formula (Eq. 6) with the modularity Q , giving place to our proposed fitness function (Eq. 7).

Starting with (a), extensive experimentation has shown that a certain transformation of the interaction index used in Eq. 2 allowed more balanced results, as detailed next. The value of Smart’s interaction index (Eq. 3) is the sum of two products of proportions (the proportion of residents in X that work in Y by the proportion of jobs in Y that are held by workers from X , plus the proportion of residents in Y that work in X by the proportion of jobs in X that are held by workers from Y). The interpretation of this interaction value is facilitated when it is transformed into a proportion-like dimension by dividing by two and computing the square root (this is similar to a geometric average):

$$I^R(X, Y) = I^R(Y, X) = \sqrt{\frac{IS(X, Y)}{2}} \quad (5)$$

This transformation of the interaction index changes the results of the Π function (Eq. 6) by shifting its optimum towards more self-contained regions, although the self-containment levels are still considerably lower than what is common practice in the field.

$$\Pi^R(\mathcal{P}) = \frac{\sum_{i \in \mathcal{P}} I^R(\{i\}, M^{\mathcal{P}, i} \setminus \{i\})}{N} \quad (6)$$

To compensate this tendency, we propose to combine Eq. 6 with modularity in a simple way (Eq. 7) and test it within the general framework we propose.

$$f(\mathcal{P}, I) = \frac{\Pi^R(\mathcal{P})}{1 - Q(\mathcal{P})} \quad (7)$$

We do not claim that this is the ideal fitness function in the context of LMAs delin-eations. Its use here is merely instrumental, and more research should be conducted on the suitability of current and other possible fitness functions for the definition of LMAs and other forms of FRs. Identifying a consensus on the more appropriate fitness function is, however, beyond the scope of this study.

4 Choosing a parameter setting algorithm

When the number of parameters to be tuned is small, an exhaustive search technique, such as grid search, can be applied easily to find the best possible set of parameters. As the number of possible sets to be tested grows dramatically with the number of parameters and with the size of their domains, stochastic (random) search techniques might be more time-efficient or effective, however, in finding the optimal solution or a better approximation. In this study, we consider one example of each type of search method: grid search (exhaustive) and genetic algorithm (a form of controlled random search). The exact methods are discarded because the number of possible solutions in many cases is so large that the evaluation of all of them is computationally infeasible.

4.1 Grid search

Grid search is a method that performs an exhaustive search through a manually specified subset of the parameter space of the regionalisation algorithm subject to a certain score function (the fitness function). It implements a regular hyper-dimensional search with a given step size that defines the grid. A nested grid search is also possible, in which the first level uses a large step size, and the following levels focus on smaller areas of the search space with smaller step size, which is a useful strategy to reduce the total computational time.

4.2 Genetic algorithm

Since their introduction (Holland, 1975; Goldberg, 1989), GAs have been widely used in many fields (Coley, 1999; Goldberg, 2006). A GA is a general-purpose method that can be implemented directly without any knowledge apart from the parameter domain and the evaluation function.

The specific scheme of the GA chosen here is as follows. At the start, a population of G_p possible parameter configurations is created by generating random values from the domain of each parameter. The method then iterates over a loop. In each iteration, G_c new solutions are produced by recombination of the solutions in the population. Each gene (parameter of the TTWA method) of each new solution has a probability G_m of being randomly mutated. At the end of each iteration, the G_p solutions with better scores in the fitness function are selected to form the population of the next generation (truncation selection). The process is stopped after a certain number G_i of consecutive iterations without finding a better solution (stagnation of the search).

Because this is an stochastic technique, the results from one execution to another can vary if the search gets trapped in a local maximum, and therefore it is advisable to perform several executions of this technique, apart from choosing a conservative configuration (G_p and G_i sufficiently large) that increases the chances of finding the global optimum.

4.2.1 Crossover operator

The crossover (or recombination) operator helps to explore the search space by combining the information of previously generated solutions. On each application, this operator chooses two parent solutions from the current population, with probability proportional to the fitness ranking of the solutions, and then forms a new one combining the information of both parents in the following way. For each parameter p , a random number $r \leftarrow U(0.0, 1.0)$ is generated, and the corresponding parameter of the new solution (child) is calculated as follows:

$$p_{child} \leftarrow r \times p_{parent1} + (1.0 - r) \times p_{parent2} \quad (8)$$

4.2.2 Mutation operator

The mutation operator is aimed at the introduction of randomness in the search process to avoid endogamy and improve diversity. It is applied to the new solution created by each recombination. Each parameter p has a probability G_m of being mutated, by adding a randomly generated value with a normal distribution centred in the current value of the parameter. This operation can be formulated as follows:

$$p \leftarrow N(p, s_p) \quad (9)$$

where s_p is the step size of the mutation for the given parameter p , to be set depending on the value domain of that parameter (analogous to the step sizes employed in the grid search, although the random nature of mutations allow for an unbounded precision level, i.e., a mutation step size $s_p = 0.5$ allows for changes greater, and lesser, than 0.5).

5 Evaluating the proposal

To explore the suitability of the proposed approach, the selection of methods/case studies was guided by two conditions: public availability of both a nation-wide commuting dataset and an exhaustive and unambiguous description of the method’s steps. These conditions led to selecting two contrasting territories and methods: (1) Sweden, with a base geography composed of 290 TUs (*kommuner*), and their national concept of LMA, the *Lokala-arbetsmarknader*, deriving from a method that requires setting 2 parameters, and (2) the UK, composed of 10,558 wards, and its national definition of LMAs, the so-called Travel-To-Work Areas, resulting from the application of a method that requires setting 4 parameters.

5.1 Commuting data

For the Swedish case study, we have selected the matrix of origin-destination commuting from the year 2010¹. With 290 TUs (*kommuner*), this dataset is indeed very small and should not suffer from computational restrictions regardless of the method employed as the parameter tuner.

For the British case study, we have employed the commuting data from the year 2001². This source includes data on commuting between 10,558 TUs (in this case *wards*), and such a number implies a combinatorial explosion of possible regionalisations as well as longer computing times to calculate each regionalisation.

In both cases, the year chosen allows the comparison of our results with the official sets of LMAs, which are updated annually in the Swedish case and every ten years in the British case.

Table 1 summarises some relevant indicators for both countries. Compared to Sweden, the base geography of UK has a finer-grained resolution, with smaller areas (and populations) and more unbalanced job ratios, with considerably lower self-containment levels.

5.2 Case Study 1: Sweden

5.2.1 Regionalisation algorithm: the *Lokala-arbetsmarknader* method

The method used by Statistics Sweden to define the national set of LMAs (Statistics Sweden, 2010) has three steps. The first step labels as centres the TUs where (i) the percentage of residents working within the TU’s boundaries—(supply-side) self-containment—exceeds 80% and where (ii) the percentage of residents commuting to

¹The commuting data for Sweden were downloaded from Sweden Statistics site (http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__AM__AM0207__AM0207L/AM0207PendlKomA04/?rxid=49c4244c-dfb4-44a9-a663-a36d92ea597d), as well as the official definitions of *Lokala-arbetsmarknader* (<http://www.scb.se/sv/Hitta-statistik/Statistik-efter-amne/Arbetsmarknad/Sysselsattning-forvarvsarbete-och-arbetsstider/Registerbaserad-arbetsmarknadsstatistik-RAMS/7899/Lokala-arbetsmarknader-LA/Forteckning-over-lokala-arbetsmarknader/>).

²The commuting data for the UK were downloaded from the UK Data Service site (<http://census.ukdataservice.ac.uk/get-data/flow-data.aspx>).

work—(supply-side) dependence—to a single destination is less than 7.5%. The second step identifies as (double-core) centres every pair of TUs where each one has the other as the destination of their largest commuting outflow, regardless of the conditions in the first step. The third step performs assignment of the rest of the TUs to the centres, through the hierarchical trees defined by the largest outflows of each non-central TU. Thus, if the destination of largest outflow of non-central A is non-central B, and the destination of largest outflow of B is centre C, A is assigned to C. With this process, all TUs become allocated to a LMA (each LMA identified by its centre)³.

The two parameters to configure in this method are the minimum self-containment (a) and the maximum dependence on a single destination (d).

5.2.2 Grid-search parameter tuning

With only two parameters to optimise, the grid search can be configured easily. We start by specifying the domains to explore for each parameter. In this case, we set broad intervals around the standard value of each parameter: $a \in [65\%, 95\%]$ and $d \in [1\%, 20\%]$. In both cases, we set the step size to 0.5%.

This resulted in 2379 applications of the LAM method with less than a second of total computation time. The best fitness value (Eq. 7), 0.372516 (for modularity 0.82145 and inner interaction 0.066512), was obtained for parameter configurations $a \times d$ with $a \in \{73\%, 73.5\%, 74\%\}$ and $d = 8\%$.

5.2.3 GA-based parameter tuning

For the GA-based search, we used the same domains and the same mutation step sizes for each parameter than the corresponding step sizes in the grid search. The rest of GA parameters were set as follows: generations without improvement $G_i = 400$, number of solutions in the initial population $G_p = 100$, probability of mutation $G_m = 0.5$, and number of recombinations (new individuals) per generation $G_c = 1$.

The GA-based search was applied 5 different times. This resulted in 3882 applications of the LAM method that took less than a second to compute. The best fitness value, 0.376556 (for modularity 0.82182 and inner interaction 0.067096), was obtained in all of the applications of the GA, for parameter configurations in $a \times d$ with $a \in [74.09\%, 74.31\%]$ and $d \in [7.86\%, 7.91\%]$.

5.2.4 Comparison

The GA-based search improved both modularity and inner interaction of the best result from our grid-search. In order for the latter to find the same result, the step size of the grid-search should be set to 0.1%, and it would evaluate 57491 different parameter configurations, causing considerably more workload than in the GA-based search. Despite this situation, both techniques can find the best possible solution in a reasonable amount of time and with similar efforts devoted to design each of the experiments.

Figure 2 shows the resulting maps for the case of Sweden. Most of the LMAs' boundaries are the same in the three maps. The main difference between the official parameter configuration and the two calculated in this work are in the divisions of some of the most populated LMAs—particularly the new LMAs of Uppsala (that stems from Stockholm) and Helsingborg (that stems from Malmö).

³Unless there are sets of three or more non-central TUs whose largest outflows create a loop within that set. However, this situation does not arise in any of the available commuting datasets for Sweden.

Table 2 summarises several statistics for the three regionalisations under consideration: the official definition (first column) and the two identified in our approach using grid search (second column) and GA (third column). These statistics includes several relevant indicators on the main characteristics that LMAs must fulfil (EUROSTAT and Coombes, 1992; Casado-Díaz and Coombes, 2011): to be self-contained (that implies having balanced supply and demand, the job ratio) and homogeneous in size, with high inner interaction and preference for higher detail (number of LMAs). The first four data rows report the values of the TTWA parameters. The rest of rows report, successively, the fitness value, the number of identified LMAs, the average inner interaction per TU as per the two indices considered (Eqs. 2 and 6), the total self-containment of the whole regionalisation (see the note on Table 1), and some quantiles on self-containment, employed residents, job ratio and area per LMA.

Most of the statistics are very similar for the three regionalisations, namely area and population quantiles. The official parameters achieve higher levels of self-containment (total and quantiles per TU) but lower values for modularity and (specially) inner interaction. The overall gain in the compound fitness function using the GA-based automatic tuning is considerable (7.9% higher). However, the appearance of a region with minimum self-containment equal to 64.47% is a concern. It seems that in this case, the chosen fitness function is slightly biased toward more divided regionalisations. This bias might be a consequence of the resolution limit in the modularity function, as noted in Section 3.

5.3 Case Study 2: United Kingdom

5.3.1 Regionalisation algorithm: the TTWA’s method

As stated in Section 1, this procedure is one of the most widely applied internationally and constitutes the official concept of LMA with a longer history of use (Casado-Díaz and Coombes, 2011). The most recent version of this method has been selected, which was used for the revision of this set of LMAs after diffusion of the data from the British Census of Population 2001 (Office for National Statistics et al., 2008). It is noticeable that this revision includes a significant reduction in the number of parameters in comparison with previous exercises, in part with the aim of reducing the arbitrariness in the decisions. The remaining parameters to tune are four: o_m and o_t are the minimum and target levels of employed residents (size) and a_m and a_t are the minimum and target levels of self-containment, which is here calculated as the minimum value of two different measures (Coombes et al., 1986): the so-called supply-side self-containment (the proportion of localised jobs in a given region held by workers who reside in that region) and demand-side self-containment (the proportion of employed residents of a given region who work within the boundaries of that region).

These four parameters are linked in a linear relationship that basically allows a trade-off between the self-containment and the size objectives, so that more populated LMAs can have a self-containment lower than the target level, and vice versa. The parameters must meet two restrictions for coherence: $0 < a_m \leq a_t \leq 1$ and $0 < o_m < o_t$ (o_m must be strictly lower than o_t to avoid divisions by zero in the validity function of the TTWA method).

5.3.2 Grid-search parameter tuning

With four parameters to optimise, the grid search becomes considerably more complex, in terms of both the time to design the experiments and the time to compute each of them, given the combinatorial explosion of possible parameter configurations. A single

application of our implementation of the TTWA algorithm to the UK territory takes approximately 1.5 seconds (depending on the parameter configuration), and so thousands of applications require hours of computation. In this case, we decided to apply a nested grid search to reduce the computational demands.

Based on the official parameter configuration applied in UK, we set the following domains to explore in the first level of the grid search, with the aim of identifying the area in the search space that encloses the best possible parameter configuration. For the self-containment parameters, we set $a_m \in [60\%, 85\%]$ and $a_t \in [65\%, 90\%]$, both with step size 2.5%. We used a different way to specify the search domain for the population size parameters by directly stating the values to check (so that step size is not used): $o_m \in 500, 1000, 2000, \dots, 8000$ and $o_t \in 1000, 2000, 4000, \dots, 64000$.

This experiment resulted in 1900 applications of the TTWA method that took 30 minutes and 49 seconds to compute. The best fitness value (0.28403) corresponded to the parameter configuration with $a_m = 0.65$, $a_t = 0.7$, $o_m = 1000$ and $o_t = 4000$. However, many other parameter configurations achieved very similar results⁴. This made the area to explore in the following grid-search level difficult to identify.

We selected 10% (190) of the parameter configurations that achieved the best fitness values. From that set, we removed the configurations for which another configuration with same values a_m and a_t and better fitness score existed. The minimum and maximum values of each parameter in the remaining selected configurations were used to decide the domains to be explored in the second search level: $a_m \in [62.5\%, 70\%]$ and $a_t \in [67.5\%, 77.5\%]$, both with step size 0.5%, and population size domains $o_m \in 500, 750, 1000, 1500, 2000$ and $o_t \in 2000, 2500, 3000, 3500, 4000, 4500, 5000$.

This resulted in 10710 applications of the TTWA method that took 5 hour, 13 minutes and 21 seconds to compute. The best fitness value (0.285243) corresponded to the configuration $a_m = 0.68$, $a_t = 0.69$, $o_m = 1000$ and $o_t = 3500$, and again local maxima with values very close to that one were found at this search level⁵. Further exploration could be performed in the domains $a_m \in 0.65, 0.69$, $a_m \in 0.68, 0.7$, $o_m = 1000$ and $o_t \in 3500, 4000$, but we decided to halt this process here, as we already knew from the GA-based search results (see below) that the best parameter configurations were already left out of the current search scope.

5.3.3 GA-based parameter tuning

As in the Swedish case, we used the same parameter domains as in the grid search. In the case of the size parameters (o_m and o_t), it was necessary to have different step sizes for each one to ensure a good performance of the procedure, given that the target size is often rather larger than the minimum size and has greater variability. The mutation step sizes were set as follows: 0.5% for a_m and a_t , 200 for o_m and 500 for o_t .

The rest of the GA parameters were set as follows: generations without improvement $G_i = 1000$, number of solutions in the initial population $G_p = 200$, probability of mutation $G_m = 0.5$, and number of recombinations (new individuals) per generation $G_c = 1$. G_i and G_p were increased with respect to the Swedish case to account for the greater size and complexity of this case.

We found that there is more sensitivity to the lower values of the parameters o_t and (more specially) o_m . This finding is consistent with the greater density of TUs in the

⁴The 23 best results had $a_m = 0.65$, $o_t = 4000$, $a_t \in [0.675, 0.9]$ and $o_m \in 500, 1000, 2000$. The 74 best results had $a_m \in 0.65, 0.675$, $o_t \in 1000, 4000$, $a_t \in [0.65, 0.9]$ and $o_m \in 500, 1000, 2000$

⁵Other representative local maxima where $a_m = 0.66$, $a_t = 0.68$, $o_m = 1500$, $o_t = 3500$ and $a_m = 0.65$, $a_t = 0.69$, $o_m = 1000$, $o_t = 3500$

lower ranks of population. Thus, we used powers of two for the initialisation of those values. The parameter configurations of the initial population were generated as follows:

$$a_m \Leftarrow U(0.6, 0.85) \quad (10a)$$

$$a_t \Leftarrow U(0.65, 0.90) \quad (10b)$$

$$o_m \Leftarrow 2^{U(9,13)} \quad (10c)$$

$$o_t \Leftarrow 2^{U(10,16)} \quad (10d)$$

To ensure that the constraints are fulfilled, if $a_t < a_m$ then both values are swapped, if $o_t < o_m$ then both values are swapped, and if $o_m = o_t$ then $o_t \Leftarrow o_t + 1$.

The GA-based search was run 5 different times. This resulted into a total of 11574 applications of the TTWA method that took 5 hours, 23 minutes and 20 seconds to compute (1 hour and 5 minutes on average). The best fitness value, 0.287738 (for modularity 0.80379 and inner interaction 0.056458), was obtained in 2 of the 5 runs, for parameter configurations (a) $a_m = 0.7037$, $a_t = 0.7504$, $o_m = 1170$, $o_t = 3700$ and (b) $a_m = 0.7036$, $a_t = 0.7652$, $o_m = 990$, $o_t = 3700$. These numbers could be rounded up to $a_m = 0.7037$, $a_t = 0.765$, $o_m = 1000$, $o_t = 3700$ and still obtain the same regionalisation, but further attempts to reduce decimal places resulted into a slightly different regionalisation.

Restricted search The official definitions of LMAs might include the need for fixing a certain value for some of the parameters embedded in the regionalisation method. An example of this situation in the TTWA case would be fixing a minimum population size for conducting robust statistical exercises, establishing a network of cost-effective employment offices, or for education planning related to active unemployment policies, among other applications.

To illustrate the effects of setting ranges or fixed values to some of the regionalisation parameters (according to policy-making or other criteria), a second exercise for UK (called “GA restricted”) was run by setting a minimum value for employed residents of 3500. We also added an additional restriction by rounding the values o_m to nearest hundred, o_t to nearest thousand and a_m and a_t to three decimal places (thus emulating the limitations of the grid-search and honouring the human preference for rounded values). The same GA parameters were used, except for the reduced parameter domains and the step mutation for o_t , which was increased from 500 to 1000 to account for the larger values of that parameter with these settings. The resulting best parameter configuration was $a_m = 0.674$, $a_t = 0.765$, $o_m = 3700$, $o_t = 7000$, with a fitness value of 0.283767.

5.3.4 Comparison

The unrestricted GA-based search greatly improved both modularity and inner interaction compared to the grid-search, which failed to locate the best area of the parameter configuration in the first search level. This could be solved by setting considerably smaller step sizes in the first search level, however that would require considerably more computational time than the GA-based search.

The reader is reminded that the chosen fitness function and its factors have not been proven to be the appropriate function to use in this context, so one cannot rely solely on those values. Table 3 reports the same statistics as in the previous case study (see Section 5.2.4).

In terms of total self-containment, LMA self-containment, job ratio balance and area homogeneity, the unrestricted GA results achieve the best scores. The main differences

with the official parameter values⁶ are that in the unrestricted GA delineation there are some smaller LMAs in terms of employed residents although with greater overall self-containment, while the biggest LMAs have become slightly larger (the maximum values of area and employed residents increase 4.1% and 16.8%, respectively). The employed residents of London’s LMA increase in the unrestricted GA-based results (from 3.85 M for official parameters to 4.5 M). Because areas have not changed considerably across all the regionalisations, we can expect similar levels of accessibility (in terms of commuting cost and time) within each LMA. The improvements in most of the relevant statistics (notably self-containment and job ratio) may be worth the loss in total inner interaction (that decreases from 0.060817 to 0.056458, a 7.7%) and the decreased number of LMAs (from 196 to 175).

The main concern when comparing the unrestricted GA regionalisation with the official one is the presence of both smaller and larger LMAs, which points to a loss in size homogeneity. When we impose a minimum value for employed residents (3500, the same as in the TTWA official definition), the proposed methodology with the chosen fitness function produces results closer to those from the official parameters. The number of LMAs remains similar (196 and 201), as well as the minimum and median figures for self-containment, area and employed residents. The maximum area remains similar to that from the unrestricted GA results although the maximum value of employed residents is now closer to the official results. The only noticeably deterioration from the quality statistics for the official results is a slight decrease in total inner interaction (1.0% for II^S and 1.8% for II^R). However, the restricted GA results improve total self-containment (+1.7%) and modularity (+1.4%).

Figure 3 shows the resulting maps for the selected parameter configurations: official, GA unrestricted and GA restricted. When the official definition of LMAs is compared to the other two, small differences and shifts in the LMAs boundaries can be identified. However, a general agreement on the main boundaries seems to arise from the three maps. The most noticeable difference can be observed in the region of London: in the unrestricted GA results, London’s LMA reaches towards the East up to Southend-on-Sea, while this coastal region belongs to a different LMA (that approximately corresponds to the county of Essex) in the other two maps. The restricted GA results are closer to the shape of official London’s LMA but still considerably bigger, especially towards the South. Because the three alternative London’s LMAs have comparable areas, we can expect similar average commuting times in the three of them, and we cannot claim that any of these alternatives are definitively better than the other.

Summing up, the proposed methodology achieves an overall improvement in the quality statistics of the results for the official (knowledge-based) parameter configuration, while the resulting regionalisations are comparable in shape and size. In this study case, between the two parameter setting techniques applied, grid-search (deterministic) and GA (stochastic), the latter manages to find better solutions while requiring less computational and design time.

It must be noted that many of the smallest areas showed in Fig. 3 correspond to TUs that are in fact part of LMAs to which they are not contiguous. This result is an acceptable one from the “raw” application of the official TTWA method, which does not impose a contiguity restriction during its process but in which a final stage of manual or automatic optimisation can be performed to properly reassign the disconnected TUs

⁶Please note that the ‘official’ set of TTWAs compared here does not correspond to the TTWA’s definition finally published (that was subjected to an unspecified series of modifications following stakeholder consultations and geographical contiguity corrections) but to the raw results from applying the TTWA with the official parameter configuration.

(which is performed as part of the process that leads to the definition of the official set of TTWAs). This final stage of contiguity restriction is not included in this work, although it could be easily considered either during the GA-based search or in a final step, as it is done in the official definition of TTWAs in the UK.

6 Discussion

Overall, the results described in the previous sub-sections support the validity of the approach proposed here. For a majority of quality indices, the resulting LMAs score better than the official sets of areas, as shown in Section 5.3.4 and, much more relevantly for the purposes of this article, the sets of parameters resulting from the empirical exercises conducted are very similar to those effectively used in the official procedures applied in UK and Sweden to define their respective national sets of LMAs. The procedure is therefore able to produce results comparable to those derived from extensive experiments guided by expert knowledge and modulated from a wide range of inputs from very diverse sources. Such experimentations involve costly and lengthy processes that in some cases include decisions that are not subject to transparent and general rules. It must be recognised, however, that these outcomes could also be an artefact of the datasets used here, at least partially⁷.

In more technical terms, it must be acknowledged that when the method to be configured has three or less parameters, and their domains can be discretised efficiently, either grid search or a stochastic technique will be able to find the best solution. However, for more parameters or larger parameter domains a stochastic technique seems to be the correct choice.

The GA-based search has shown to be capable of finding the best solutions, in terms of the available quality indices, for both study cases. The stochastic nature of the GA-based search has not been a problem: with little effort on configuring its own parameters, few repetitions of the procedure allow for an efficient (compared to grid-search) search where the best known result is the most repeated one. Nevertheless, it must be noted that an increase in the number of repetitions and the number of iterations without improvement (the termination condition) should be considered in any specific territory with a greater number of TUs (the main source of complexity in this problem) to improve the possibilities of identifying the optimum parameter configuration and ensure an accept-

⁷This point refers primarily to the appropriateness of the combined fitness function proposed. To further explore this matter we conducted additional tests in which we used a GA to calibrate the TTWA method using the alternative fitness functions applied to the cases of the US (Census 2000) and Spain (Census 2001). Overall, the results confirm our previous findings (detailed in Section 3): the combined fitness function proposed in Section 3.3 achieves results that are a reasonable compromise between the levels of cohesion and self-containment in the identified LMAs, compared to those deriving from the application of the fitness functions previously used. Firstly, in the Spanish case the interaction-based fitness functions yield excessively low minimum values for the self-containment parameter resulting in a high atomisation of the LMAs identified. The function based on modularity yields results in the other extreme, and produces a macro-region around Madrid that is at odds with what is commonly assumed to be the Madrid labour market. The combined fitness function produces a set of much more balanced LMAs. Secondly, the US case produces different results. Although the parameters estimated using the interaction functions can be considered to be in the standard range of values, they result in an excessive fragmentation (e.g. some of the biggest metropolitan areas such as New York are divided into diverse LMAs). The function based on modularity leads to values over 90% for the self-containment parameter and the associated set of LMAs includes a relevant number of excessively large regions (10% of LMAs had area > 41,000 km²). The best configuration from the combined fitness function results in a higher number of LMAs with much more acceptable median and average sizes and where no metropolitan areas are fragmented. Further details on this additional tests are available from the authors on request.

able stability in the results. This can be done in a robust manner by applying statistical techniques to detect the stagnation (convergence) of the search process as in Trautmann et al. (2009). Among the possible improvements of the search performance are performing a fine-tuning of the GA parameters (namely the parameter domains, the mutation step sizes, the stop condition and the number of repetitions) as well as using adaptive mutation operators (Lobo et al., 2007) and including local optimisers as in a memetic algorithm (see a review in Moscato and Cotta, 2010).

7 Conclusions

The decision on the appropriate values for the parameters typically embedded in the methods used for the definition of functional regions, such as local labour market areas, in many countries is a critical step that usually relies on the knowledge held by the experts that conduct such exercises. They normally apply a trial and error procedure through which parameters are applied and the associated alternative geographies contrasted, until a set of areas is assumed to reflect well the underlying phenomena and the image of the functional reality which is tacitly shared by the different actors involved in these processes. However, the resulting scale and the specific set of boundaries chosen exert a crucial influence on the many different policy making dimensions in which these geographies are used in the countries where they are defined, which makes this decision critical.

In this study, we propose an approach that supports the quantitative calibration of the methods used to define LMAs and illustrate it through its application to case studies of sufficiently large dimensions and contrasting features so that the results are examples of real-world applications. In the specific illustration of the approach developed here, two alternative procedures (a genetic algorithm and a generic grid search) are used as “wrappers” to set the parameters involved in the application of two well-known instances presented in the literature: the methods developed in the UK and Sweden for the definition of their national sets of LMAs. Both regionalisation methods are different in nature (e.g., the Swedish method departs from the identification of foci while the British one does not) and also in the number and characteristics of the parameters that must be set.

The overall evidence has shown that the parameter sets resulting from the proposed approach produce LMA configurations that score better than the official delineations of LMAs in terms of the two indices most used in the literature. A significant feature is that in all cases the parameter values identified are relatively close to those used in the official methods. This similarity is very relevant for our purpose as it shows the usefulness of the method, if it is assumed that the existing sets of official LMAs configurations are the fittest for the territories and functional relations under consideration. However, it seems fair to recognise that more research is probably needed to incontrovertibly assure the robustness of the approach, since these results could also be in part an artefact of the datasets specifically used in the empirical analyses conducted.

The range of potential uses of the method proposed here is wide, with the cautions already outlined. It could be a useful tool for assisting official delineations’ updates following the availability of new data or the revision of the criteria used and, more significantly, a support for starting-up such processes in countries or regions that are new to this concept. The choice of a specific set of LMAs against its alternatives has significant implications in statistical and policy-making terms. For this reason, a particularly relevant context for this approach would be the definition of LMAs in countries with a federal

structure (where regional interests could be in conflict) or to undertake international exercises such as an European level cross-national definition of LMAs. The latter is an example of a potentially controversial case since (a) European countries are very diverse in terms of spatial units and commuting patterns, (b) many countries have their own methods and parameter sets, which have historically proven to work well for them, and (c) other countries do not have any experience. In this context, expert knowledge would face a very complex decision derived from the wide diversity in terms of territorial reality and previous experience. Fixing a ‘technical’ set of parameters that fits such a complex geography as a departing point for further discussions, in the course of which this initial set could be modified to introduce many other relevant dimensions, could undoubtedly contribute to the success of the process. Moreover, the empirical exercises that have been performed in this study include an example of how the approach is flexible and can adapt to different situations. One such circumstance would be the need for fixing a minimum value for one of the parameters, as would happen if, for example, there was a need for setting a minimum population size for the identified LMAs due to statistical sampling requirements or employment offices location planning.

It must be noted that our approach does not intend to replace the experts or policy makers in their task of identifying suitable parameters (that eventually determine the number and shape of the LMAs) but rather to support their decision-making process by offering acceptable combinations in terms of the desirable conditions that these geographies should meet assuming that they can be codified in logical/mathematical terms and take the form of a fitness function.

As stated before, the experimentation section of this study has shown how this methodology can address large problems (as is the case of the UK). The application of this approach to even larger instances such as the whole set of EU countries (something that has not been attempted until now) might imply days of computation time. However, such running times should not be an obstacle in a context where obtaining high-quality solutions is the main priority. Moreover this drawback could be easily alleviated through the use of parallel computation and the allocation of more computing resources, as the techniques applied are easily parallelisable.

The method has been validated in two case studies: the UK and Sweden, that understand LMAs based on different criteria. This shows that the proposed methodology is easily adaptable to any spatial regionalisation technique and concept (such as housing markets and transportation areas) used either in the administrative or the academic spheres, by adapting the fitness function accordingly.

Future work should focus on the study of the available fitness functions for spatial functional regionalisation and eventually the development of a more accurate quality measure for this concept. Additional lines of work could test the possibility of improving the GA-based parameter-setting approach in four ways: (i) the fine-tuning of the parameters of the GA algorithm itself, (ii) the adjustment of the selection strategies (notably the truncation scheme selection, which could be a source of premature stagnation caused by loss of diversity), (iii) the introduction of a local optimisation strategy to reduce the time needed to reach the final solution and, when applied to the specific problem whose solution has been illustrated here (the definition of LMAs), and (iv) the performance of a comparative analysis of alternative fitness functions.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (grant numbers CSO2011-29943-C03-02 and CSO2014-55780-C3-2-P, National R&D&i Plan). The funding source had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the paper for publication.

References

- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008: P10008
- Cantú-Paz E (2004) Feature subset selection, class separability, and genetic algorithms. In: *Genetic and Evolutionary Computation—GECCO 2004*, 959–970. Springer
- Casado-Díaz JM, Coombes M (2011) The delineation of 21st century local labour market areas: A critical review and a research agenda. *Boletín de la Asociación de Geógrafos Españoles* 57: 7–32 & 393–396
- Casado Yusta S (2009) Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters* 30: 525–534
- Coley DA (1999) *An introduction to genetic algorithms for scientists and engineers*. World Scientific
- Coombes MG, Green AE, Openshaw S (1986) An Efficient Algorithm to Generate Official Statistical Reporting Areas: The Case of the 1984 Travel-to-Work Areas Revision in Britain. *Journal of the Operational Research Society* 37: 943–953
- Cörvers F, Hensen M, Bongaerts D (2009) Delimitation and coherence of functional and administrative regions. *Regional Studies* 43: 19–31
- EUROSTAT, Coombes MG (1992) étude sur les zones d’emploi. Technical Report E/LOC/20, Office for Official Publications of the European Communities, Luxembourg
- Farmer CJ, Fotheringham AS (2012) Network-based functional regions. *Environment and Planning A* 43: 2723–2741
- Flórez-Revuelta F, Casado-Díaz JM, Martínez-Bernabeu L (2008) An evolutionary approach to the delineation of functional areas based on travel-to-work flows. *International Journal of Automation and Computing* 1: 10–21
- Fortunato S, Barthelemy M (2007) Resolution limit in community detection. *Proceedings of the National Academy of Sciences* 104: 36–41
- Fusco G, Cagliioni M (2011) Hierarchical clustering through spatial interaction data. the case of commuting flows in South-Eastern France. In: *Computational Science and Its Applications-ICCSA 2011*, 135–151. Springer
- Goldberg DE (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional
- Goldberg DE (2006) *Genetic algorithms*. Pearson Education India
- Holland JH (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, USA
- ISTAT (1997) *I sistemi locali del lavoro 1991*. Istituto Nazionale di Statistica (ISTAT), Roma

- John GH, Kohavi R, Pfleger K (1994) Irrelevant features and the subset selection problem. In: *Machine Learning: Proceedings of the Eleventh International*, 121–129. Morgan Kaufmann
- Lancichinetti A, Fortunato S (2011) Limits of modularity maximization in community detection. *Physical Review E* 84: 066122
- Leicht EA, Newman ME (2008) Community structure in directed networks. *Physical review letters* 100: 118703
- Lobo F, Lima C, Michalewicz Z (2007) *Parameter Setting in Evolutionary Algorithms*, volume 54. Springer Verlag
- Martínez-Bernabeu L, Flórez-Revuelta F, Casado-Díaz JM (2012) Grouping genetic operators for the delineation of functional areas based on spatial interaction. *Expert Systems with Applications* 39: 6754–6766
- Moscato P, Cotta C (2010) A modern introduction to memetic algorithms. In: Gendreau M, Potvin JY (eds.) *Handbook of Metaheuristics, International Series in Operations Research and Management Science*, volume 146, 141–183. Springer US
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Physical Review E* 69: 026113
- OECD (2002) Redefining territories: The functional regions. Technical report, Organisation for Economic Co-operation and Development (OECD), Paris
- Office for National Statistics, Coombes MG, Bond S (2008) Travel-to-work areas: the 2007 review. Technical report, Office for National Statistics (ONS), London
- Office of Management and Budget (2000) Standards for defining metropolitan and micropolitan statistical areas. Federal Register / Vol. 65, No. 249 / December 27, 2000 / Notices / pp. 82227–82238
- Saeyns Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics* 23: 2507–2517
- Smart MW (1974) Labour market areas: uses and definition. *Progress in Planning* 2: 238–353
- Statistics Sweden (2010) Construction and use of labour market areas in sweden. Technical report, Statistics Sweden, Örebro
- Trautmann H, Wagner T, Naujoks B, Preuss M, Mehnen J (2009) Statistical methods for convergence detection of multi-objective evolutionary algorithms. *Evolutionary computation* 17: 493–509

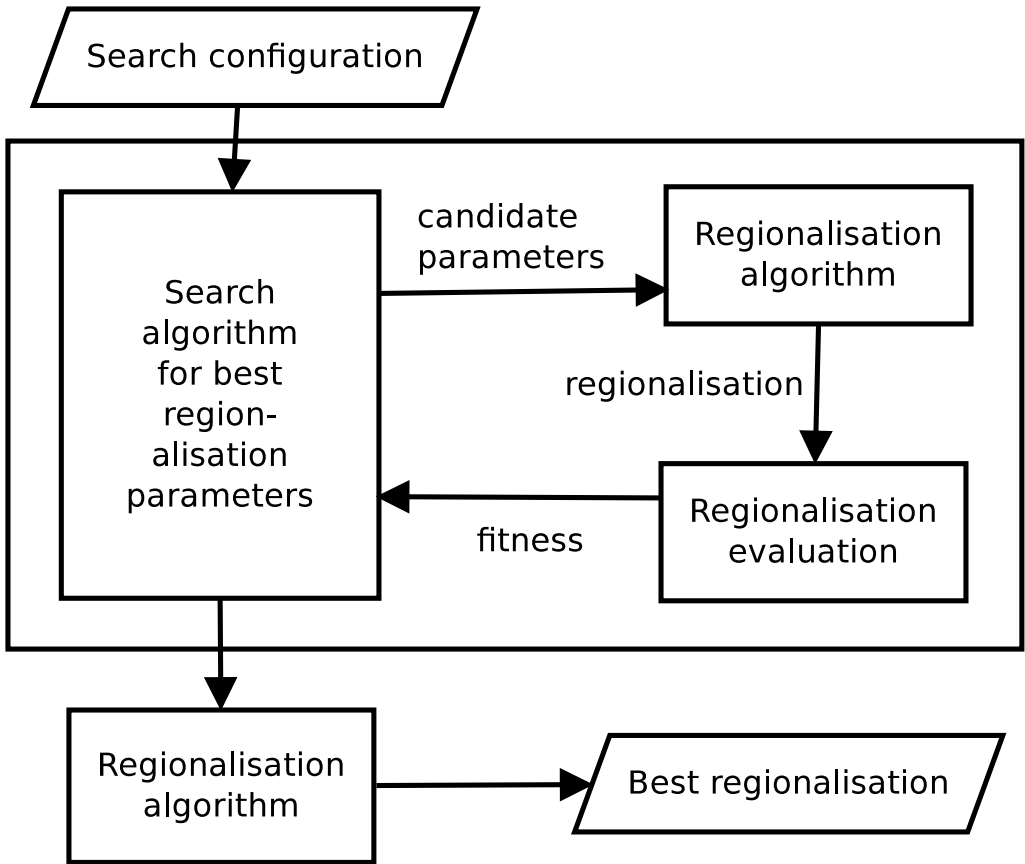


Figure 1: Wrapper model for regionalisation algorithms

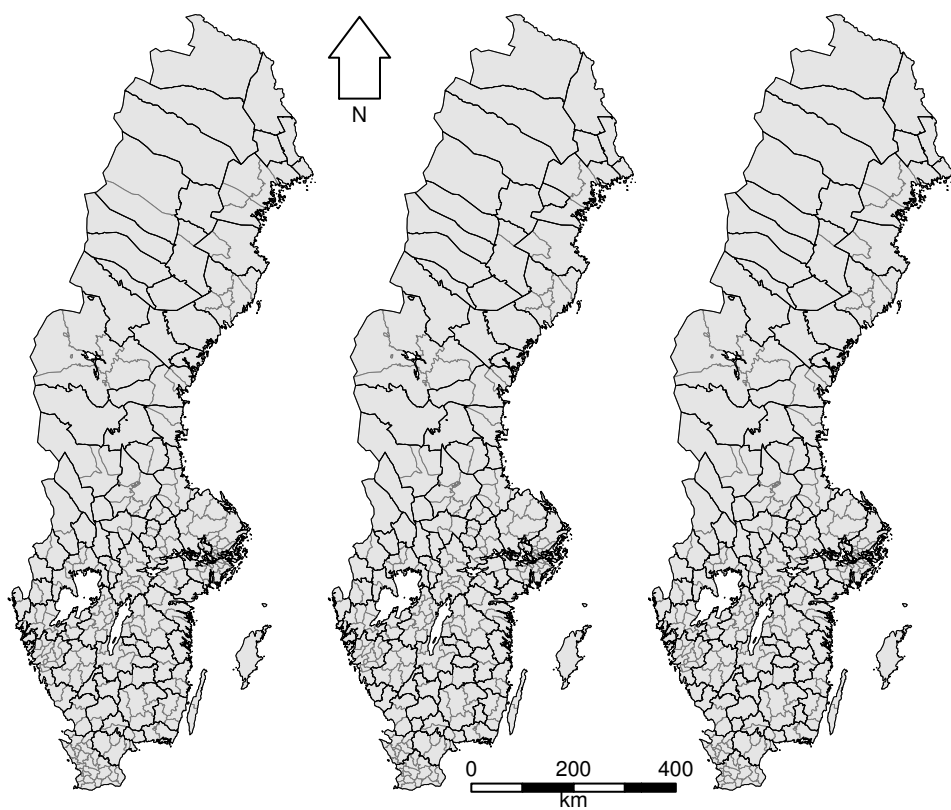


Figure 2: Maps for Sweden 2010 LMAs: official (bottom), GS (centre) and GA (top)

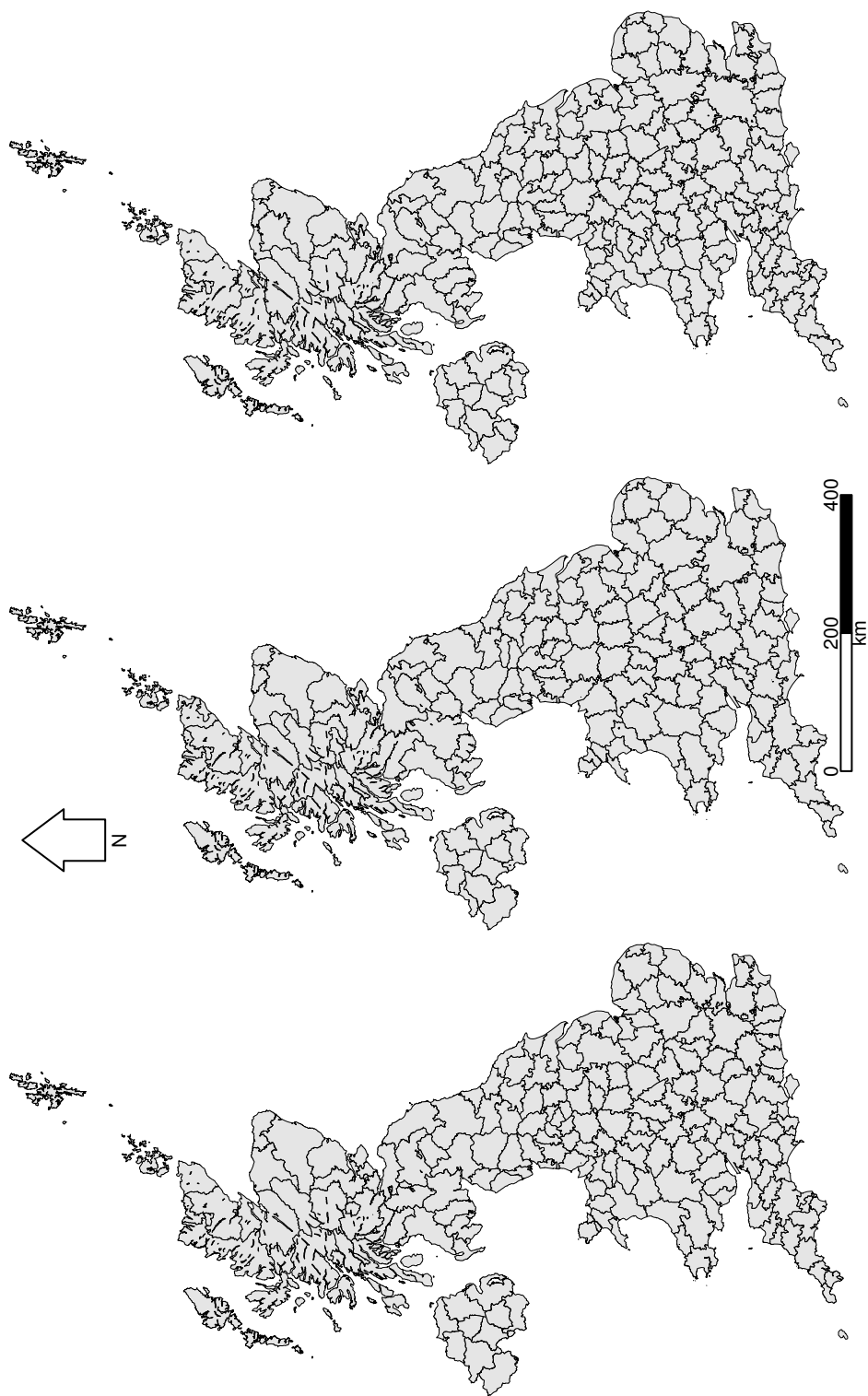


Figure 3: Maps for UK 2001 LMAs: official (bottom), GA (centre) and GA restricted (top)

Table 1: Characteristics of the territorial units for both case studies

	Sweden (2010)	UK (2001)
Territorial units	290	10,558
TU's Employed residents		
- Minimum	1,007	237
- Low decile	2,794	879
- Median	7,067	2,032
- Maximum	434,508	17,725
TU's job ratio		
- Minimum	0.405	0.098
- Median	0.909	0.668
- Maximum	1.937	247.163
TU's area (Km ²)		
- Minimum	8.71	0.13
- Low decile	150.13	1.17
- Median	673.17	4.86
- Maximum	19,371.12	3,320.72
TU's self-containment		
- Minimum	13.44%	0.21%
- Low decile	32.62%	12.54%
- Median	64.29%	20.16%
Total Self-Containment	67.43%	23.75%
Modularity (Eq. 4)	0.6513	0.2373

Source: own calculations. Notes: The job ratio is the number of jobs divided by employed residents of an area. Self-containment represents the minimum between two measures: supply-side and demand-side self-containment, as defined in Section 1 (one value for each TU). Total Self-Containment is defined as the proportion of persons who reside and work in the same TU over the national figure of employed residents (one value for the whole territory considered).

Table 2: Parameter estimates and statistics for best obtained results (Sweden 2010)

	Official	Grid-search	GA
Parameter estimates			
Min. core self-containment a	80.0%	74.0%	74.3%
Max. core dependence d	7.5%	8.0%	7.9%
Statistics of the regionalisations			
Number of LMAs	76	89	87
Inner interaction			
- II^S (Eq. 2)	0.0060	0.0066	0.0066
- II^R (Eq. 6)	0.0647	0.0665	0.0671
Modularity (Eq. 4)	0.8145	0.8215	0.8218
Fitness value (Eq. 7)	0.3490	0.3725	0.3766
LMA's self-containment			
- Minimum	75.95%	64.47%	64.47%
- Low decile	80.42%	78.27%	79.01%
- Median	85.90%	84.72%	85.18%
Total self-containment	92.31%	90.71%	90.75%
LMA's employed residents			
- Minimum	1,245	1,211	1,211
- Low decile	2,713	2,775	2,732
- Median	16,768	15,632	15,997
- Maximum	1,197,405	1,050,563	1,050,563
LMA's job ratio			
- Minimum	0.886	0.876	0.886
- Low decile	0.943	0.922	0.930
- Median	0.980	0.978	0.978
- Top decile	1.027	1.043	1.042
- Maximum	1.117	1.171	1.171
LMA's area (Km ²)			
- Minimum	883.23	139.18	139.18
- Low decile	1,356.91	1,072.56	1,068.74
- Median	3,929.85	3,151.44	3,209.80
- Maximum	27,410.03	27,410.03	27,410.03

Notes: see notes in Table 1.

Table 3: Parameter estimates and statistics for UK 2001 best results

	Official	Grid-search	GA	GA restric.
Parameter estimates				
Min. self-containment a_m	66.67%	68.00%	70.37%	67.40%
Target self-containment a_t	75.00%	69.00%	76.50%	73.90%
Min. residents o_m	3,500	1,000	1,000	3,700
Target residents o_t	25,000	3500	3,700	7,000
Statistics of the regionalisations				
Number of LMAs	196	211	175	201
Inner interaction				
- II^S (Eq. 2)	0.0051	0.0051	0.0046	0.0051
- II^R (Eq. 6)	0.0608	0.0598	0.0565	0.0597
Modularity (Eq. 4)	0.7781	0.7904	0.8038	0.7893
Fitness value (Eq. 7)	0.2741	0.28524	0.28774	0.28357
LMA's self-containment				
- Minimum	66.76%	68.02%	70.39%	67.46%
- Low decile	68.56%	69.59%	72.09%	68.55%
- Median	77.29%	77.82%	80.52%	77.38%
Total self-containment	80.94%	82.63%	84.55%	82.62%
LMA's employed residents				
- Minimum	3,769	1,267	1,267	4,792
- Low decile	11,889	4,649	4,781	9,589
- Median	64,105	50,599	58,316	52,261
- Maximum	3,850,565	4,077,467	4,499,072	4,157,887
LMA's job ratio				
- Minimum	0.798	0.810	0.835	0.810
- Low decile	0.888	0.878	0.900	0.882
- Median	0.967	0.969	0.969	0.964
- Top decile	1.034	1.029	1.027	1.034
- Maximum	1.130	1.134	1.126	1.129
LMA's area (Km ²)				
- Minimum	151.50	107.71	107.71	145.48
- Low decile	406.75	419.43	581.09	414.49
- Median	1,036.74	986.27	1,204.19	1,004.91
- Maximum	5,061.20	5,061.20	5,271.94	5,281.88

Notes: see notes in Table 1.